and land cover classes. Owing to the very heterogeneous distribution of rain (Section 5.2), the northern half of the domain is obviously much wetter than the southern half at the end of the two-week period.

The value of the observations for retrieving the uncertain initial condition is obvious. Without assimilating the brightness data, our best guess for the top node saturation is the prior solution shown in the middle row of Figure 6.3, which is quite far from the truth. Once we assimilate the brightness data, we are able to accurately estimate the initial condition. With the initial condition being the dominant source of uncertainty in this experiment, we also get good estimates of the entire saturation time series.

Figure 6.4 shows time series of the area average errors in the top node saturation and in the soil temperature. The errors shown are root-mean-square errors (rmse) of the prior and the estimated fields with respect to the true fields. In the legend we also indicate the time average of the area average rmse's. Note that the soil moisture errors are in terms of the relative saturation, which varies between zero and one. In order to derive the errors in terms of volumetric moisture percent, the numbers have to be scaled with the porosity. For reference, the area average porosity is 0.46.

It is obvious that on average the estimated saturation fields are much closer to the truth than the prior fields. This again confirms that there is enough information in the brightness observations to infer the true initial condition for the saturation. Another interesting feature of Figure 6.4 is how the error of the estimate varies with time. In particular, the error decreases each time it rains. In other words, precipitation events tend to wipe out the memory of the system including errors.

Note that the errors in the soil temperature estimates are unrealistically small, because the experiment was not designed to test the soil temperature estimate thoroughly. In brief, the soil temperature initial condition was assumed perfectly known, and there was only a small uncertainty in the forcing of the soil temperature equation. These constraints lead to an overly optimistic estimate of the soil temperature if compared to field conditions.

It is important to point out why the area average prior saturation error in Figure 6.4 decreases with time. This is due to the specific setup of Reference Experiment I. In this particular setup, the initial condition becomes unimportant after some time because of the nonlinearity of the system. If, for instance, we start from too wet an initial saturation, then the evapotranspiration will also be higher than in the true case, therefore reducing the difference between the prior and the true fields with time. Conversely, if we start too dry, then evapotranspiration will be suppressed, and the prior will again edge towards the truth. Since in this experiment we use the same model and the same micro-meteorologic forcings for the generation of the (synthetic) true field and for the estimates, and since we only have a small model error, we know that the prior, the estimate, and the true fields must finally converge.

If this setup were to hold true in nature, we could indeed reasonably well estimate the saturation at the final time without a complex assimilation algorithm, in fact we could even do without the brightness data. In reality, however, we will of course never know the model physics and the forcings well enough to rely on this naive idea. Nature will always be much more complicated than our models, and we will need to estimate the initial condition over and over again. The important point for the interpretation of Reference Experiment I is therefore to look at how well we are doing in estimating the initial condition. Clearly, we can much improve the prior trajectories and we can estimate the initial condition to satisfying accuracy if we assimilate the brightness observations.

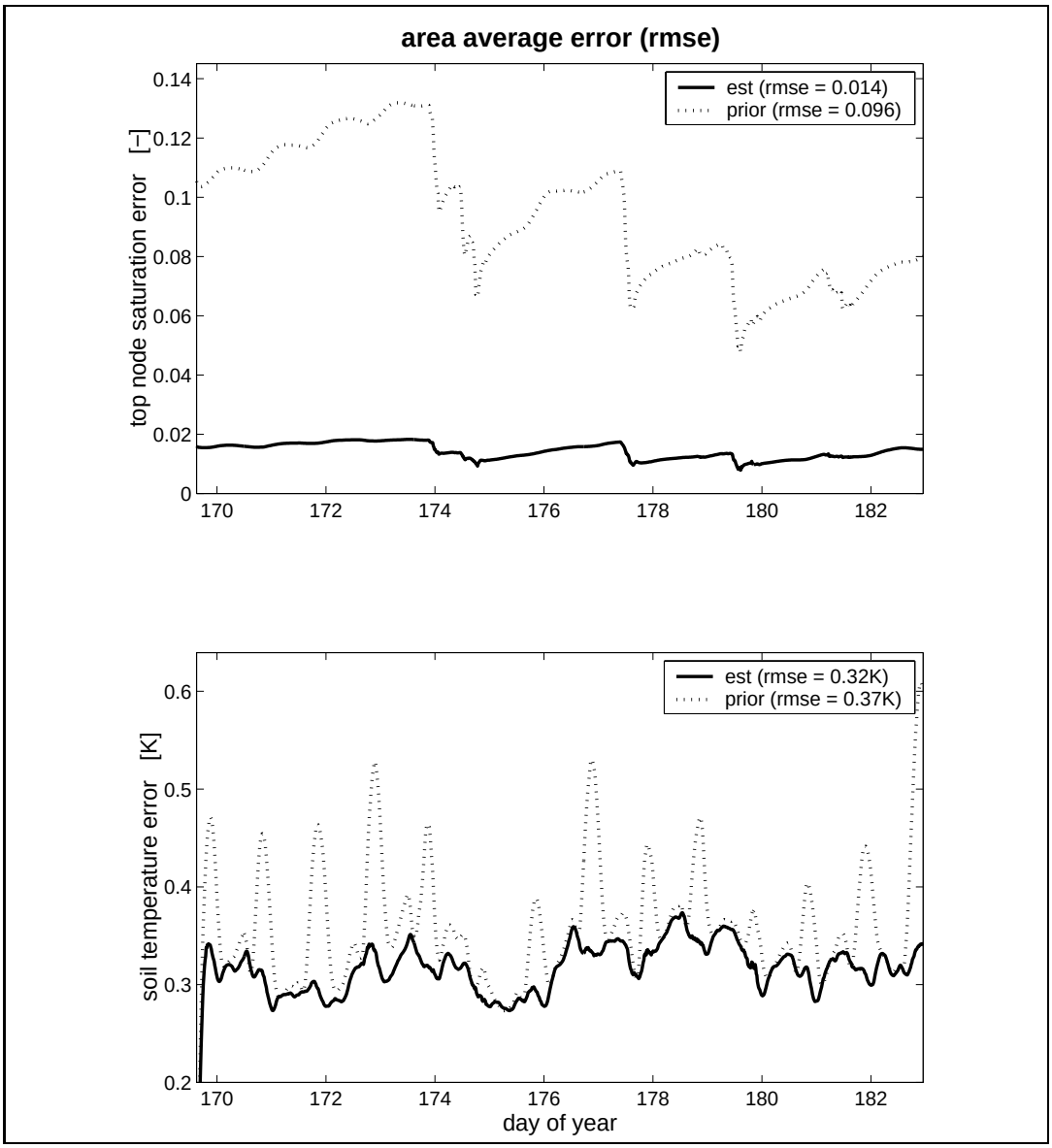

Figure 6.4: Area average errors for Reference Experiment I. The root-mean-square errors (rmse) of the prior and the estimated top node saturation and soil temperature with respect to the (synthetic) true fields are shown. In the legend we also indicate the temporal average of the area average rmse. Note that the soil moisture errors are in terms of saturation. To derive the errors in terms of volumetric moisture percent, the numbers have to be scaled with the porosity. The area average porosity is 0.46. Obviously, the assimilation greatly improves the errors in the top node saturation over the prior errors. Note that this experiment is not designed to test the soil temperature estimate. Also note that the decrease in the prior rmse is an artefact of the setup of Reference Experiment I (see text).

6.1.3 Profile Estimation and Validity of the Land Surface Model

For all subsurface nodes, we get the same excellent estimates as for the surface node (see Figure 6.14 for Reference Experiment II in Section 6.2.1). However, the fact that the profile estimates are this good does not in itself mean that the profile information is necessarily contained in the brightness observations. Indeed, recall that the shape of the initial condition profile is fixed as hydrostatic in both the generation of the true solution and in the assimilation. Moreover, in our ideal setup we use the same hydrologic model for the generation and for the estimation of the true fields. Finally, by using only a small model error in the top flux boundary condition, we assume an almost perfect model. For these reasons we naturally get excellent results for the profile estimates.

Whether subsurface information can in fact be retrieved from measurements related to the states in the top few centimeters is mostly determined by the accuracy and the physical realism of the land surface model. Without assimilating measurements that are directly related to the subsurface states, the only way information can be propagated to the deeper soil is via the hydrologic model. However, the time scales for the evolution of the profile are longer than the intervals that could reasonably be covered by a single assimilation window. Consequently, experiments to show the benefit of brightness assimilation for the estimation of the saturation profile depend on the successful development of an operational framework for soil moisture data assimilation.

Experiments to test the validity of the profile estimates could be verified with independent observations of the subsurface states. But the meaning of the large-scale (surface and subsurface) saturation is itself not obvious, considering that each pixel covers an area the size of tens of square kilometers. Certainly one cannot expect to go out to the field and verify the estimates with point measurements of the profile saturation. The large-scale saturation profile may rather be understood as an aggregate measure of how much water is stored at depth across the pixel. Such a vertically distributed and nonlinear reservoir may be used to determine how the land surface interacts with the atmosphere. In other words, the reservoir may tell us how much large-scale evapotranspiration and how much large-scale infiltration can be sustained by the land surface.

But suppose that we assimilate brightness data for a long time in an operational fashion. After enough time has elapsed, the posterior data residuals will show whether the subsurface dynamics of the hydrologic model and the prior statistics are consistent with the data. If this is the case, we have proof that the land surface model captures the multilayer subsurface reservoir and its feedback onto the surface states with reasonable accuracy. We can then interpret the saturation profiles in the manner suggested. If, on the other hand, the posterior data residuals remain biased or have structure even after months of assimilation, the model turns out not to be adequate in its description of land surface dynamics. In this case, we would have learned something about land surface dynamics. The assimilation then prompts us to go back and improve the hydrologic model accordingly.

6.1.4 Reduced Objective Function

Variational data assimilation is based on minimizing an objective function. Consequently, it is important to closely examine the value of the objective. Figure 6.5 shows the prior and the reduced objective as a function of the iteration on the tangent-linearization. In the first iteration, the objective function decreases significantly from its prior value of 31,749 to 6705. Unlike in a gradient search, the objective function need not decrease monotonically

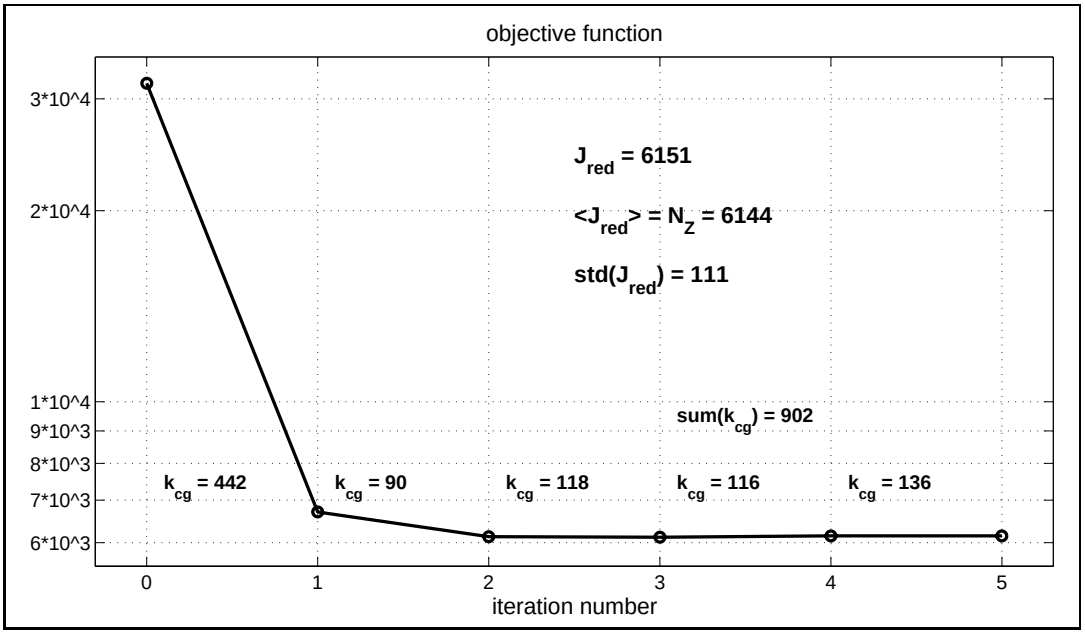

Figure 6.5: Objective function versus iteration number for Reference Experiment I. The reduced objective function after convergence is 6151. The number of data points is 6144, which is also the expected value of the reduced objective function. The standard deviation of the reduced objective function is 111. The values of k_{cg} indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

with the number of iterations. In fact, upon closer inspection we see that the reduced objective equals 6132 after the second iteration, 6120 after the third iteration, 6150 after the fourth iteration, and finally 6151 after convergence. During the first iteration, the linearized trajectory is adjusted to a first-cut estimate using all observations. Had the problem been linear from the outset, we would have been done after this first iteration. During the remaining iterations, the assimilation algorithm mainly ensures that the estimate obeys the nonlinear state equation by making small corrections to the estimated trajectory. In other words, during the early iterations the estimates are not yet dynamically consistent, and the value of the objective function may decrease or grow while dynamic consistency is achieved during the iteration process.

Note that the final reduced objective function $\hat{J} = 6151$ compares well with the expected value of $\overline{J} = 6144$, which is equal to the number of data points (Section 2.3.6). The standard deviation of the objective function in this case is $\sigma_J = 111$. Since the errors were generated synthetically with perfectly known covariances, and since the true fields were generated with the same model that is used in the inversion, the reduced objective must by design indicate consistency of the prior assumptions with the (synthetic) data. This is an important validation step in the practical implementation of the algorithm.

6.1.5 Posterior Data Residuals

After the estimates have been derived, it is advisable to also take a close look at the posterior data residuals. The data residuals are the difference between the estimates of the measured quantities and the observations (Equation (2.32), Section 2.4). For an optimal estimate, the residuals must not show spatial or temporal patterns, that is they ought to be white in space and time. Moreover, their distribution should be close to normal with a mean of zero. If all of the above holds true, we can assume with reasonable confidence that the estimation process was optimal.

First, we examine the mean values (Figure 6.6). For all of the twelve individual residual brightness images that have been assimilated, we find a mean whose 95% confidence interval includes zero. But the mean for all residuals with a 95% confidence interval is $0.14 \pm 0.12K$, which does not include zero. Indeed, Figure 6.6 suggests that there is a slight positive bias. The fact that the sample mean of all residuals is not compatible at 5% significance level with the hypothesis of a zero mean could just be a spurious statistical fluctuation of the given realization. But the effect is more likely to have its origin in the nonlinearities of the hydrologic model. We will return to this point below when we test for normality. Note that the residuals of Reference Experiment II do not show this bias (Section 6.2).

Before further investigating the posterior data residuals, it is best to standardize them. Ideally, this is done with the posterior error covariance of the data residuals (2.37). However, computational limitations prohibit the exact calculation of this quantity. We therefore normalize with the sample standard deviation of the data residuals. The sample standard deviation of the residuals of the individual images varies between $4.4K$ and $5.1K$. The sample standard deviation for all residuals is $4.8K$. Recall that the measurement error standard deviation is $5K$.

Figure 6.7 shows the posterior data residuals for all twelve brightness images that have been assimilated in the experiment. The residuals of each image are standardized with the sample mean and standard deviation of the corresponding observation time. The residuals show no obvious spatial structure, which indicates that the estimation algorithm works

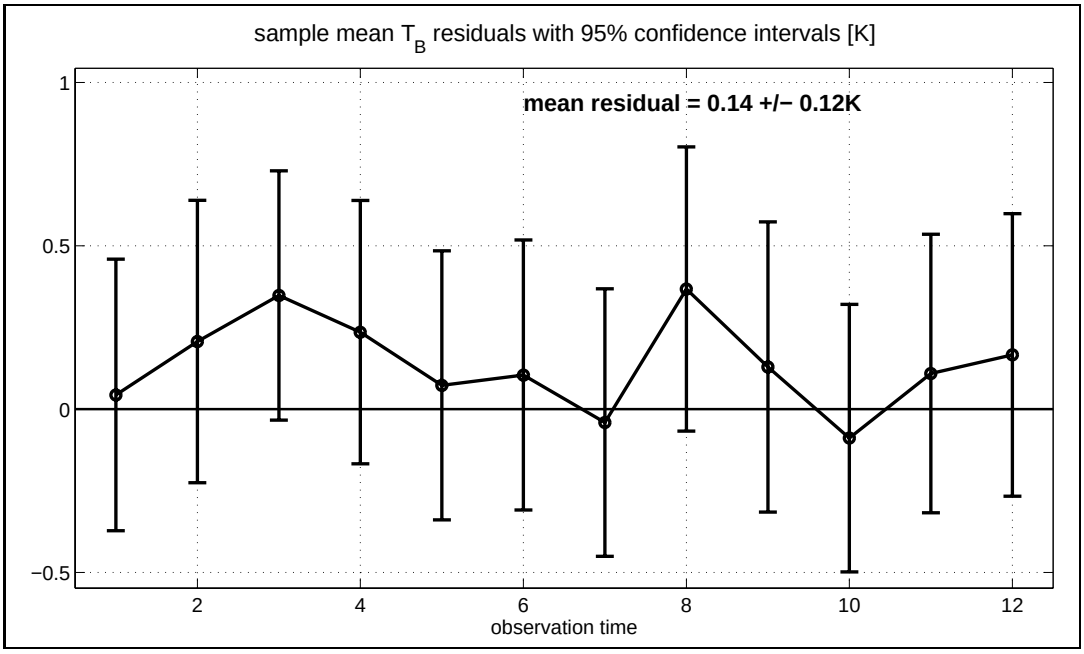

Figure 6.6: Sample mean values of the posterior data residuals for Reference Experiment I. The mean residuals for the twelve images that have been assimilated are shown together with the 95% confidence intervals. Even though the confidence interval for each of the images includes zero, the confidence interval for the sample mean of all images does not. This is likely due to the nonlinearities in the hydrologic model.

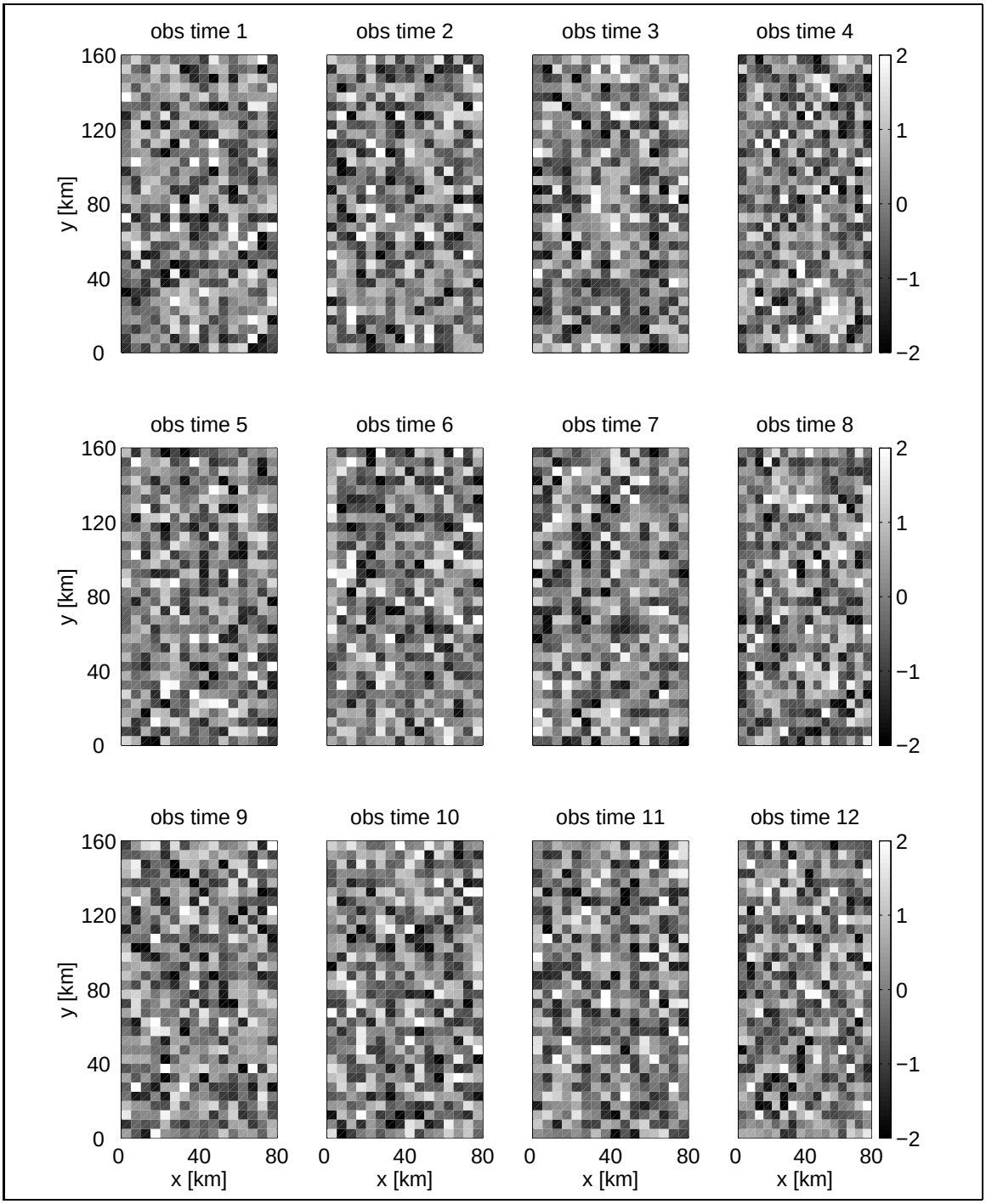

Figure 6.7: Standardized posterior data residuals for Reference Experiment I for the twelve brightness images that have been assimilated. The residuals of each image are standardized with the sample mean and standard deviation of the corresponding observation time. The residuals show no obvious spatial structure, which indicates that the estimation algorithm works optimally.

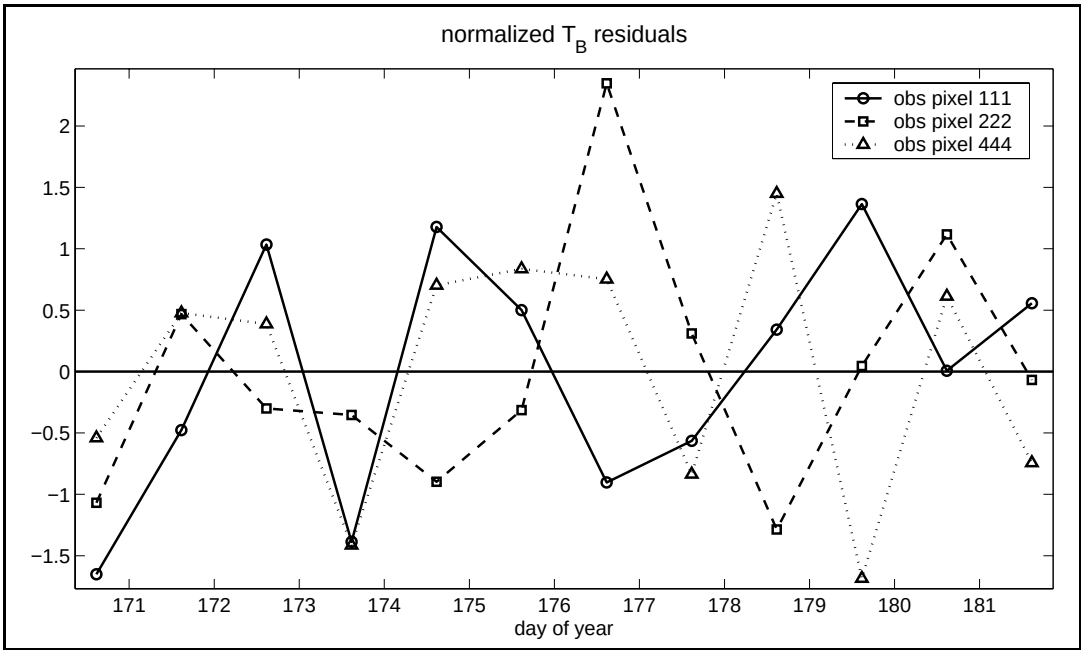

Figure 6.8: Standardized posterior data residuals for Reference Experiment I at three observation pixels from different areas within the domain. The residuals have no obvious temporal correlation, which indicates that the estimation algorithm works optimally.

optimally. Likewise, Figure 6.8 shows the time series of the data residuals for three pixels from different regions of the domain. The posterior data residuals also show no obvious temporal correlation. A test for whiteness using the autocorrelation function [Jenkins and Watts, 1968] confirms this result, but it must be noted that the short length of the time series defies an accurate statistical assessment.

Finally, Figure 6.9 shows the sample cumulative distribution functions (cdf) of the standardized posterior data residuals for two brightness images. For comparison, we also show the theoretical cumulative distribution function of the standard normal distribution. We can test the posterior residuals for normality with the Kolmogorov-Smirnov test [Benjamin and Cornell, 1970]. In brief, the Kolmogorov-Smirnov test compares the maximum difference between the sample cdf and the assumed theoretical cdf.

For two out of the twelve observation times, we must reject the hypothesis of a normal distribution of the data residuals at a 5% significance level. The upper panel of Figure 6.9 shows the sample cdf at the second observation time, for which we must reject the normality hypothesis. In contrast, the lower panel shows the sample cdf of the residuals at the tenth observation time. Here, the hypothesis of a normal distribution is not inconsistent with the data. Since the model and the statistics used for the generation of the (synthetic) true fields and for the estimation are identical in this experiment, and since the assimilation algorithm used is optimal in the linear case, we must attribute the deviation from normality to the nonlinear effects. Given the strong nonlinearities in the land surface model and in the measurement operator, it is not surprising that the posterior data residuals are not perfectly normally distributed.